DOCUMENT RESUME

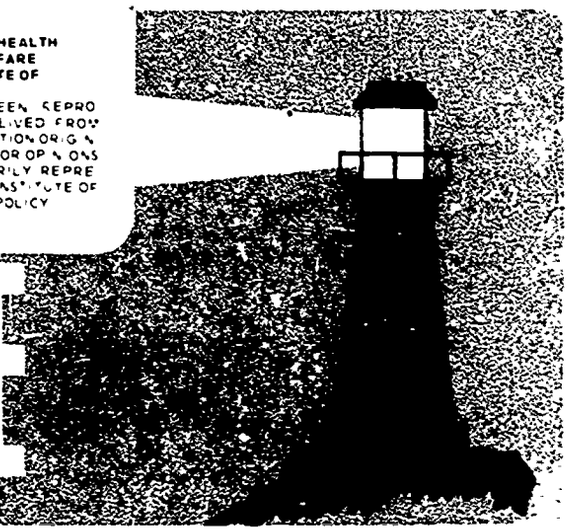
ED 078 084 TM 002 903

AUTHOR	McCluskey, Lawrence
TITLE	Program Evaluation: An Overview.
INSTITUTION	Metropolitan School Study Council, New York, N.Y.
PUB DATE	Feb 73
NOTE	6p.
AVAILABLE FROM	Metropolitan School Study Council 525 West 120th Street, New York, N.Y. 10027 ($5.00 per year)
JOURNAL CIT	MSSC Exchange; v31 n6 p1-5 Feb 1973
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Educational Objectives; Educational Programs; *Evaluation Methods; Learning Processes; *Models; Program Development; *Program Evaluation; Publications; Systems Approach

ABSTRACT

Various models of educational evaluation are presented. These include: (1) the classical type model, which contains the following guidelines: formulate objectives, classify objectives, define objectives in behavioral terms, suggest situations in which achievement of objectives will be shown, develop or select appraisal techniques, and gather and interpret performance data; (2) the accreditation model, which emphasizes the process of education, rather than its outcomes; (3) the systems model, inherent in which is the idea of evaluation as a management feedback system throughout the course of the program; and (4) the discrepancy model, which combines the best available methods for using evaluation as a program development tool. (CK)

MSSC EXCHANGE



Volume XXXI, Number 6 February 1973

Program Evaluation: An Overview

by Lawrence McCluskey

Evaluation became an educational issue when the pattern of governmental funding of education underwent a dramatic change in the years following the 1957 launching of Russia's Sputnik satellite. The National Defense Education Act (1958) and the Elementary and Secondary Education Act (1965) caused federal support of education alone to almost double. But most of the programs funded under these new auspices were geared toward innovation and change, rather than at expanding or enhancing educational techniques that were previously in existence. As a result of this emphasis on "innovative" programs, governmental agencies began demanding some kind of monitoring progress by which their effectiveness could be gauged. The old axiom, "Why throw good money after bad?" gained new currency, and "evaluation" became a new career route for many educational researchers.

However, in education, as well as in allied fields, there is an important difference between "evaluative" research and what is often called "pure" research. The difference is simply this: Where "pure" research asks the question, 'Is treatment "a" a suitable remedy for deficiency "b"?' evaluative research assumes that the answer to this question is yes and proceeds to examine the impact of the treatment on the indicated deficiency. In other words, "pure" research begins with hypotheses, while "evaluative" research begins with assumptions. Or, stating the concept in a different way, the "pure" researcher cannot be "wrong;" his hypothesis is either accepted or rejected according to some predetermined standard. On the other hand, the "evaluative" researcher is faced with determining whether or not some assumption is "right or wrong." At least, this was the state of affairs in the earlier models used in evaluating educational programs.

The Classical Type Model

One of the first models used in evaluation studies, which is in fact sometimes referred to as the Classical Type of Evaluation, contains the following guidelines:

"1. **Formulate objectives.** Determine broad goals of the program.
2. **Classify objectives.** Develop a typology of objectives so an economy of thought and action may be achieved.
3. **Define objectives in behavioral terms . . .**
4. **Suggest situations in which achievement of objectives will be shown.**
5. **Develop or select appraisal techniques.** (standardized tests, ad hoc tests, questionnaires, etc.)
6. **Gather and interpret performance data.** The final step in the evaluation process involves the measurement of student performance data with behaviorally stated objectives..." [1]

Even a cursory examination of this model reveals what has now been recognized as one of its weaknesses — its emphasis on examining program products or outcomes. In effect, it not only assumes the efficacy of some activity or treatment, but also assumes the presence of the activity in an effort to produce the objectives formulated by the designers of the program. But experience has shown that the chief impediment in the implementation of a new education program is often the failure to properly apply the treatments specified in the program design.

Furthermore, this model assumes that all educational objectives can be measured by objective, quantifiable methods. Now this assumption might be supported (given some operational definition of attainment) if educational objectives were confined to those areas in which standardized testing has been established, but it encounters serious objections when one considers such program objectives as "improved student citizenship" or "enhancing appreciation of individual worth." Obviously, these points do not imply that verifiable performance outcomes are not the concern of the evaluators. In fact, verification that the program has attained its objectives is the focal point of evaluative research. However, if the evaluator concentrates solely on the achievement of behaviorally defined performance objectives as this model dictates, he may overlook those aspects of the program which greatly influence its success or failure.

The Accreditation Model

Another method of program evaluation is what may be called the Accreditation Model. In this model, emphasis is on the process of education, rather than on the outcomes. The assumption made in this model is that improvement in the educational process would result in improvement in producing desirable outcomes. Criteria were developed for rating various components of the education process such as building facilities, library size and services, instruction equipment, teacher qualifications, guidance programs, etc. Once these criteria were established, a team of experts would visit the program site and rate the program on various criteria. The ratings of these experts could then be used to compare one program to another, or to a set of standards laid down by other experts in the field of education.

Opinion is divided over the usefulness of the accreditation model as an evaluation

---

ABOUT THE AUTHOR Lawrence McCluskey has served as a Research Associate with the Field Evaluation Unit of the Horace Mann-Lincoln Institute and is currently the Associate Director of the Institute of Administrative Research at Teachers College Columbia University

technique. Its advantages appear to be that it can quickly respond to the need for program evaluation and makes use of the abilities of people who are "experts" in their fields. These advantages, however, may be offset by some weaknesses which are inherent in the model.

Some observers have commented that although the accreditation model "has the advantages of quick response and the utilization of the full range of the evaluator's competence, it obviously leaves much to be desired in terms of objectivity and validity, which are at best moot."[2] The "genetic defects" in the Accreditation Model are "that its practitioners do not seek to justify empirically the standards used to judge worth and that attention to the processes of education is not balanced by attention to its consequences on learners."[3]

The Systems Model

Still another type of evaluation utilizes a Systems Approach. Inherent in this model is the notion of evaluation as a management feedback system throughout the course of the program. Evaluation in this context becomes a monitoring process which concentrates on gathering data about programs and providing management with information necessary to make modification and improvement during the course of the program. A summary of the way the systems model operates is included here.[4]

A **specific example** of an evaluation design based on the Systems Approach is one developed by Stufflebeam called the CIPP model.[5] Each letter stands for a discrete step in evaluation: Context, Input, Process, Product; together, they can be applied to nearly any educational evaluation study. Each part, as Suchman suggested, has its own objectives.

The major objective of **context evaluation** is to define the environment where change is to occur, the environment's unmet needs, problems underlying those needs, and opportunities. Information from **context evaluation** is ultimately used to establish program goals and objectives.

The purpose of an **input evaluation** is to determine how to utilize resources to meet the program goals and objectives. The end product of such an evaluation is an analysis of alternative procedural designs in cost/benefit terms, from which the decision maker can select. Decisions based upon input evaluation usually result in the specification of procedures, schedule, staff requirements, and budget. According to Stufflebeam, by evaluating the input, it can be decided whether other types of inputs are needed to achieve the objectives.

Once a designed course of action has been approved and implementation of the design has begun, **process evaluation** is needed to provide periodic feedback to project managers and others responsible for continuous control and refinement of plans and procedures. The objective of **process evaluation** is to detect or predict, during the implementation stages, defects in the procedural design or its implementation.

Finally, **product evaluation** is used to determine the effectiveness of the project after it has run full cycle. Its objective is to relate outcomes to objectives, and to context and input, i.e., to measure and interpret outcomes.

The CIPP model, while offering a helpful theoretical frame of reference for the assessment of change has been found to be deficient as a guide for actual practice. The complexity of its analysis of evaluation into many decision-making situations has made it unmanageable except in theory. "In short, while the proposed structure (CIPP) provides a general guide for developing evaluation designs, educators must still engage heavily in the laborious, painstaking process of developing each design **de novo**."[6]

The Discrepancy Model

The dearth of readily applicable theory, and the virtual absence of reported successful evaluation practice with a systems orientation was fully recognized by the evaluation team headed by Malcolm Provus, when they set out to construct a new model. "Our mandate was clear: Redefine the purpose of evaluation . . . and then devise and test an operational evaluation model based on sound theory . . ." (p. 2). The **Discrepancy Evaluation Model 1969**, known as the Provus Report, seems to combine the best available methods for using evaluation as a program development tool as well as a means of program assessment, in a readily adaptable, workable model.

In developing the Discrepancy Model, Provus proceeded on the following basic assumptions (among others):

"1. Many educational programs . . . are installed in public school systems without adequate planning.
2. Given this fact, evaluation should be a process for program development and stabilization, as well as a means of assessment. To accomplish this purpose, evaluation must provide information which decision makers can use to improve, stabilize and assess programs. (p. 8-9).

Provus sees evaluation, at its simplest level, as the comparison of performance against a standard. Like Stufflebeam, he divides the evaluation process into stages. In each stage, some indicator of performance is obtained which is compared to a standard which serves as the criterion of performance. The relationship among different evaluation stages, and between performance and standard at each stage, are illustrated schematically in Figure I.

An educational program is viewed as a dynamic input-output system with specifications for inputs, process and output being necessary and sufficient for program design. The relationship among these components may be represented by the following equation:

$$I\,(P) - 0$$

where "I" - input, "P" - process, and "O" - output. "Outputs" are viewed as a function of the interaction of inputs with process. For example, students, teachers and materials (inputs) interact in such a manner (process) as to produce a change in reading levels (output). The difference between the "goal" of the program and the "output" of the program should be minimized for program success. (p. 4).

Evaluation Stages

Stage	Performance	Standard
I	Program Design Input Dimension Process Dimension Output Dimension	Design Criteria
II	Program Operation	Program Design Input Dimension Process Dimension
III	Program Interim Products	Program Design Process Dimension Output Dimension
IV	Program Terminal Products	Program Design Output Dimension
V	Program Cost	Cost of Other Programs with Same Product

FIGURE I

In Figure I, the "standard" at stage I is the Design Criteria — a comprehensive list of program elements that make up the three basic "systems" categories of input, process and output. The Design Criteria (a sample of which appears on page 17 of the Provus Report) constitutes a basic assumption on which all other criteria for standards used throughout the evaluation are based. Provus believes it is vital to the smooth operation and ultimate success of any program that these Design Criteria be formed with information provided by program staff, and preferably in a "design meeting" attended by these people.

When the Design Criteria have been agreed upon, then, a description of the program's design is obtained as "performance" information. Stage I evaluation takes place when the program design is compared with the Design Criteria. Discrepancy between "performance" and "standard" is reported to those responsible for management of the program. To eliminate the discrepancy and approach congruence between the two, adjustments may be made in either one.

Once the program design has been established, the program is ready to be implemented, and the "standard" against which the "performance" of initial implementation is measured becomes that program design (Stage II). Once again, discrepancy information provided by the evaluator may be used by the program manager to redefine the program or change installation procedures.

Not until Stage III is any cause and effect comparison made. At Stage III, the "standard" is that part of the program design which describes the relationship between program processes and interim products. Discrepancy information is used either to redefine process and relationship of process to interim product or to better control the process being used in the field.

At Stage IV the "standard" is that part of the program design which refers to terminal objectives. Program "performance" information consists of criterion measures used to estimate the terminal effects of the project.

Finally, at Stage IV, a cost benefit analysis may be done to determine program efficiency, using the cost of other programs with the same product as a standard.[7]

Defining Objectives

At this point, it is clear that if evaluation is to be a meaningful endeavor it must specify the degree to which a program has achieved its objectives. But this consideration raises a further question that is frequently encountered in the field situation. Many programs, especially those funded by state and federal agencies, tend to be composed of multiple objectives, some of which are clearly specified in the program outline, while others are only hinted at. Examination of many program proposals reveals that the designers, in an effort to get the maximum advantage out of the program, have failed to delineate between or among various types of objectives. As a result, evaluation of such programs becomes an impossible task, and both planners and evaluators find themselves frustrated in their labors.

It appears then that great care should be taken in specifying program objectives according to some coherent plan. One such plan was developed by Operation PEP in California a few years ago. According to this paradigm objectives are divided into four stages; policy objectives, program objectives, curriculum objectives and instructional objectives. Each of these phases in the chain of objectives is assigned a person or group who would be accountable for the objectives at that particular level.

Operation PEP objectives are defined as follows:

1. **Policy objectives** define the performance commitments of an organization; they define ends that must be achieved to fulfill external (societal) and internal (organizational) requirements. Objectives at this level would normally be associated with the policymaking body, the Board of Education.
2. **Program objectives** are derived from policy objectives; they define a plan of action for the achievement of the internal and external purposes stated in the policy objectives. Program objectives would be the responsibility of the program director or coordinator, or possibly assistant or district superintendent.
3. **Curriculum objectives** would define the performance outcomes required to fulfill the **program** requirements. Administrators would be accountable for objectives at this level.
4. **Instructional objectives** refer to performance in the actual teaching-learning process. They refer to individual and instructional staff performance products, and are usually in the hands of teachers.[8]

Examination of this model shows that the more generalized the specified objective, the higher in the organizational chain responsibility for its attainment lies. Conversely, those objectives which are most specifically stated are the responsibility of the people who are in most immediate contact with the population impacted by the program. Such an organization allows an evaluator not only to recognize a program dysfunction, but also to trace the dysfunction to its source, and to feedback necessary information to program directors while there is still time to correct the dysfunction. Stating this same proposition in another way, whereas the classical model of evaluation would only permit an evaluator to state that a specific objective had not been achieved, this model would allow him to point out the reason that it was not achieved. For example, suppose that a board of education decided that teaching machines would enhance the level of student achievement (Policy level) and decided to purchase such machines for use in an individual school (Program level). The local school administrator (Curriculum level) could scarcely be held accountable for

attainment of the final objective, enhanced student achievement, (Instructional level) if the machines were never delivered. In addition, evaluators working with such a model would be able to provide information about this specific dysfunction to the policy level body during the course of the program so that the discrepancy between planned outcome and actual conditions could be reasonably reduced.

One More Evaluation Paradigm

In an attempt to draw upon the research that has been discussed here and to develop an evaluation model adaptable to the widest variety of program designs, a number of IAR staff members reviewed many program designs and proposals and attempted to apply various evaluation models to these designs. A second step in this process was to compare the original program designs with final evaluation reports. This was done in an effort to determine the degree to which the programs had achieved their stated objectives as well as the degree to which evaluators might have used their findings to assist those responsible for implementation of the program. As a result of this survey, of documents, the staff attempted to develop a new evaluation model. (See Figure II).

Program Assumption:	(Develops from needs assessment/ treatment identification)
Policy Statement:	(Generalization)
Program Goal:	(Specific performance desired)
Curriculum Objective:	(Selection of one or more alternative means/methods/sequences to achieve desired performance)
Resources Necessary and Allocated:	(Acquisition of staff, facilities, materials, etc.)
Conversion:	(Training, orientation, scheduling)
Processes:	(Instruction, performance)
How Performance is to be Assessed:	(Impact analysis)

FIGURE II

Obviously this model derives many of its features from models that have been previously discussed. but it also contains two other significant features: Resource Allocation and Conversion. While Resource Allocation is a self-explanatory term, the notion of Conversion merits elaboration here.

Conversion, a notion found to be critical in the program planning evaluation interface, deals with what occurs **after** various program inputs (personnel, supplies, facilities, etc.) have been allocated and **before** the process begins. Let me be more specific. If we think about any innovative program that is to be introduced into a system, we soon realize that is illusory to imagine that the staff, who are unfamiliar with the new techniques and facilities, not to mention the students, will immediately begin to function at their optimum levels. Rather, experience, as well as common sense, dictates that there should be a time of training and adjustment built into the time sequencing of the program, and that information on the progress of the program during this time be supplied to the program planners. The necessity of this process is recognized in the model in the step called Conversion.

Perhaps the use of this evaluation model might be clarified by actually applying it to a single program feature. To go back to the example cited previously, let us suppose that a district wished to measure the effectiveness of using teaching machines to improve student achievement. The completed model might resemble **Figure 3.**

Program Assumption:	Highly structured, self-pacing instruction that includes immediate feedback will improve student learning.
Policy Statement:	Teaching machines will be used as part of the instructional process in secondary school mathematics
Program Goal:	To use teaching machines in instruction of general mathematics to eighth grade boys in school X beginning with the spring semester.
Curriculum Objective:	To significantly improve achievement in math in two classes of eighth grade students by using teaching machines in the classroom for a period of one hour per day.
Resources Allocated and Necessary:	Teaching machines, programmed text, teacher, classroom, students.
Conversion:	Training of staff by central office math supervisor and familiarization of students by local teachers with the techniques necessary to properly use teaching machines.
Processes:	Daily use of teaching machines for a one hour period in two fifth grade math classes.
Performance Assessment:	Statement of testing strategy and sequence. Analysis of test results.

FIGURE III

This IAR model represents an attempt to achieve a number of ends. First, it tries to insure concruency of thought between planner and evaluator from generalized policy statement to specific end desired. In practice, participation in the program design should involve an array of staff members and district officers. The program structure should help to make clear the roles of the various participants in implementation. Secondly, it provides a means by which the evaluator can monitor the program throughout its duration and provide feedback to those responsible for program implementation. Furthermore, it also allows causes of program dysfunction to be quickly isolated and discrepancies between specification and performance to be reduced. Finally, it makes explicit the means by which achievement of program objectives can be measured.

FOOTNOTES

1 Glass, Gene V "The Growth of Evaluation Methodology," presented at the Evaluation Workshop of the Adult Education Research Conference, Minneapolis, Minnesota, March, 1970 p 13

2 Guba, Egon G and Stufflebeam, Daniel L **Evaluation: The Process of Stimulating, Aiding, and Abetting Insightful Action** An address delivered at the Second National Symposium for Professors of Educational Research, Boulder, Colorado, November 21, 1969

3 Glass, Gene V **op. cit.** p 22

4 The discussion of the CIPP and the Discrepancy Model is taken from a memo written by Bruce Dollar, who was a Research Associate with the Field Evaluation Unit of the Horace Mann-Lincoln Institute during the 1970-71 school year

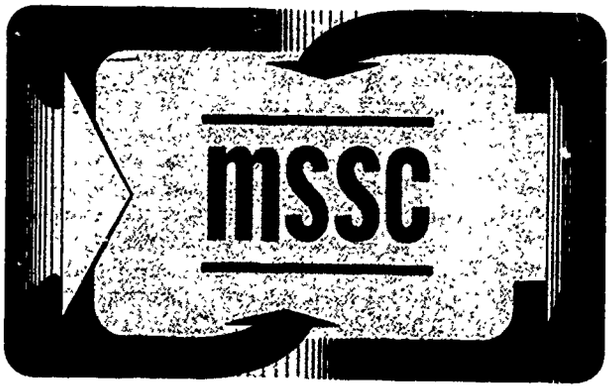
5 Stufflebeam, Daniel L "Toward a Science of Educational Evaluation" **Educational Technology.** Vol VIII, No 14, July 30, 1968.

6 Worthen, Blaine "Toward a Taxonomy of Evaluation Designs" **Educational Technology.** Vol VIII, No. 15, August 15, 1968

7 Provus, Malcolm **Discrepancy Evaluation Model 1969** Pittsburgh Public Schools, Pittsburgh, Pennsylvania 1969

8 Miller, Donald et al **A Manager's Guide to Objectives** Operation PEP, A State-Wide Project to Prepare Educational Planners for California, Los Angeles, California October, 1969

Second-class postage paid at New York, New York and at additional mailing offices.



METROPOLITAN SCHOOL STUDY COUNCIL

Exchange is published monthly, September through June, by METROPOLITAN SCHOOL STUDY COUNCIL an affiliate of the

INSTITUTE OF ADMINISTRATIVE RESEARCH
TEACHERS COLLEGE – COLUMBIA UNIVERSITY
525 West 120th Street, New York, New York 10027
Volume XXXI, Number 6 February 1973